PCT/DK 2004/000630

REC'D 08 OCT 2004

WIPO PCT

Kongeriget Danmark

Patent application No.: PA 2003 01356

Date of filing: 18 September 2003

Applicant: (Name and address) Nuevolution A/S
Rønnegade 8, 5.
DK-2100 København Ø
Denmark

Title: A method for obtaining structural information concerning an encoded molecule

IPC: -

This is to certify that the attached documents are exact copies of the above mentioned patent application as originally filed.

Patent- og Varemærkestyrelsen
Økonomi- og Erhvervsministeriet

01 October 2004

Susanne Morsing

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

PATENT- OG VAREMÆRKESTYRELSEN

Modtaget

18 SEP. 2003

PVS

Title

A method for obtaining structural information concerning an encoded molecule

Technical Field of the Invention

The present invention relates to the method for obtaining structural information about an encoded molecule. The encoded molecule is usually produced by a process that comprises the reaction of a plurality of chemical entities. The synthesis of the encoded molecule is recorded or programmed in an identifier oligonucleotide which is attached to the encoded molecule. The structural information obtained by the present method may be used to obtain the entire structure of the encoded molecule or a part thereof.

Background of the invention

The generation of molecules carrying new properties remains a challenging task. Recently, and number of procedures have been suggested that should allow a more efficient generation and screening of a huge number of molecules. The approach taken may involve the encoding and/or templating of molecules other than natural biopolymers and a coupling of the molecules to respective templates or identifier parts containing information about the reactants that have participated in the formation of the molecule. These approaches allow the researcher to generate and screen a huge number of molecules at the same time.

In US 5,723,598 it is suggested to prepare libraries of bifunctional molecules, in which one part of the bifunctional molecule comprises an encoded part and the other part of the molecule contains an identifying part. The identifying part is segregated into codons, i.e. a stretches of nucleotides, which codes for reactants that have been involved in the synthesis of the encoded molecule. The libraries of bifunctional molecules are generally prepared by a split-and-mix method, which involves the initial reaction between a nascent bifunctional molecule and a range of different reactants in separate compartments at one

end of the nascent bifunctional molecule and a corresponding range of identifier unit oligonucleotides (codons) and the other end. Subsequently, the contents of the compartments are mixed and the mixture is disposed in separate compartments and reacted again with another range of reactants and corresponding codons. Following the generation of a library of the bifunctional molecules, a partitioning with respect to affinity towards a target is conducted and the identifier part of the bifunctional molecule is decoded to establish the chemical structure of the compounds in the library that is likely to be a ligand to the target. The decoding step implies that the identifier oligonucleotides initially are amplified by PCR. The PCR product is subsequently incorporated in to a suitable vector which is transformed to a host organism, usually E. coli. Following the incubation of the E. coli, colonies are picked and sequenced.

Halpin and Harbury have in WO 00/23458 suggested an improvement to the approach stipulated above. The approach is based on the same split-and-mix strategy for synthesis of combinatorial libraries comprising two or more synthetic steps. A plurality nucleic acid templates are used, each having at one end a chemical reactive site and dispersed throughout the stand a plurality of codon regions, each of said codon regions in turn specifying different codons. Separately, each of the strands, identified by a first codon region, is reacted at the chemical reaction sites with specific selected reagents. Subsequently, all the strands are pooled and subjected to a second partitioning based on a second codon region. The split-and-combine method is conducted an appropriate number of times to produce a library of typically between 10^3 and 10^6 different compounds. The decoding is performed utilizing the process depicted above.

Recently, new a method for encoding molecules has been suggested, which can be performed in a single "pot". WO 02/00419 and WO 02/103008 disclose methods for preparing virtually any molecule connected to a template coding for chemical entities which have reacted to form the molecule. In

short, a template segregated into a plurality of codons and a plurality of building blocks comprising a transferable chemical entity and an anticodon are initially provided. Under hybridisation conditions, the template and building blocks are annealed together and the chemical entities are subsequently reacted to form the molecule. However, after a sufficient number of rounds of selections have been performed, the template must be decoded to establish the identity of the encoded molecule. The decoding step implies that the template oligonucleotides initially are amplified by PCR. The PCR product is subsequently incorporated in to a suitable vector which is transformed to a host organism, usually E. coli. Following the incubation of the E. coli, colonies are picked and sequenced.

In an aspect of the invention, it is the object to facilitate the decoding of the coding oligonucleotide in order to obtain at least partial structural information of the encoded molecule being a ligand towards a target. In another aspect of the invention, it is desired to obtain information about which chemical entities that result in encoded molecules successful in a selection process. Such chemical entities may be used in the formation of a second generation library.

Summary of the Invention

The present invention concerns a method for obtaining structural information about an encoded molecule produced by a process comprising reaction of a plurality of chemical entities, said encoded molecule being capable of forming part of a complex also comprising an identifier oligonucleotide containing codons informative of the identity of chemical entities which have participated in the formation of the encoded molecule, the method comprises mixing a primer oligonucleotide with the identifier oligonucleotide, subjecting the mixture to a condition allowing for an extension reaction to occur when the primer is sufficient complementary to a part of the identifier oligonucleotide, and evaluating, based on measurement of the extension reaction, the presence, absence, or relative abundance of one or more codons.

The method according to the invention may be performed on a single identifier oligonucleotide or a composition of identifier oligonucleotides to obtain structural information about the encoded molecule or a composition of encoded molecules, respectively, that have been attached to the identifier oligonucleotide(s).

A single identifier may be analysed using the above method to verify the incorporation into the encoded molecule of one or more chemical entities or to deconvolute the identity of the entire encoded molecule. A composition of two or more identifier oligonucleotides generally results from a selection process, i.e. a process involving subjecting a library of different complexes to a condition partitioning the composition from the remainder of the library. Usually the partitioning condition includes an affinity assay in which the library of complexes is contacted with a target and the identifier oligonucleotides of the binding complexes are harvested.

The conditions allowing for an extension reaction to occur may be selected from a enzymatic or chemical means. Suitably, the condition involves one or more enzymes. In a certain embodiment of the invention, the condition which allows for an extension reaction to occur includes a polymerase or a ligase as well as suitable substrates for the enzyme used. Preferred is a polymerase together with a blend of (deoxy)ribonucleotide triphosphates. Suitably, the blend include one or more of dATP, dGTP, dCTP, and dTTP.

A library of complexes can have any appropriate size. Typically, the size is above 10^3, typically above 10^6 different complexes. An effective, extensive, and rapid decoding is therefore desirable. The method of the present invention may be used at various stages of the process of finding a ligand to a certain target. As examples, the method of the invention may be used for controlling the quality of a starting library. The information acquired may be used to verify which codons being present, absent, and, in some embodiments,

also the relative abundance. Thus, the method of the invention delivers a reliable picture of the process which has produced the library. If, for some reason, a chemical entity has not been incorporated into the encoded molecules, the absence of a codon for this chemical entity will in certain embodiments of the invention indicate this fact.

Another example of the use of the present method is following the selection. After the selection has been performed the codon profile is indicative of the chemical entities that have been used in the synthesis of encoded molecules having an affinity towards the target. In the event the selection has been sufficient effective it may be possible directly to deduce a part or the entire structure of binding encoded molecules. Alternatively, it may be possible to deduce a structural unit appearing more frequently among the encoded molecules after the selection, which gives important information to the structure-activity-relationship (SAR). If the selection process has not narrowed the size of the library to a manageable number, the formation of a second generation library may be contemplated. In the formation of the second generation library chemical entities which have not been involved in the synthesis of encoded molecules that have been successful in the selection may be omitted, thus limiting the size of the new library and at the same time increasing the concentration of binding complexes. The second generation library may then be subjected to more stringent selection conditions to allow only the encoded molecules with a higher affinity to bind to the target. The second generation library may also be spiked with certain chemical entities suspected of increasing the performance of the final encoded molecule. The indication of certain successful chemical entities may be obtained from the SAR. The use in a second generation library of chemical entities, which have proved to be interesting for further investigation in a preceding library, may thus entail a shuffling with new chemical entities that may focus the second generation library in a certain desired direction.

The relative abundance of codons may make it possible to decode a plurality of identifiers simultaneously, even in the case when two or more identifiers contain the same codon. Thus, following the formation and selection of a first, second or further generation library, the identity of binding encoded molecules may be partly or entirely deconvoluted by the present method.

In a practical approach the library comprises complexes with identifier oligonucleotides having n codon positions each. In a certain aspect of the invention n is an integer independently selected from of 2 to 8. It may be preferred that n is constant among all the complexes in the library to facilitate the decoding process. Each of the codons in a certain position is in an aspect of the invention selected from a set of m different codons. m may vary for each codon position or may be constant among the various codon positions. It may be preferred in some embodiments to have all the codons in each position selected from the same set of m codons. However, in other embodiments, especially such involving hybridisation in the recognition between the codon and the anticodon, it may be preferred that all the codons are different.

Any member of the codon set differs from any other codons in the set with the identity of at least one nucleotide, i.e. at least one nucleotide position occurs. In some aspects of the invention it is preferred that any member of the codon set differs with at least two nucleotides nucleotide positions from any other member of the set to increase the fidelity of the method. In general, it is desired to maximize the differences between individual codons of the set. In some embodiments of the invention, a set of primers comprising a sequence of complementing the set of codons are prepared.

In a preferred aspect of the method a framing sequence is related to each of the n codon positions in a particular complex, said framing sequence positions the reaction of a chemical entity in the synthesis history of the encoded molecule. Typically, the framing sequence is identical among the complexes for each of the reaction rounds and is selected from a group of n different

nucleotide sequences. In a certain aspect of the invention n x m different primers fully or in part complementing any combination of the set of m different codons and the set of n different framing sequences is prepared. The n x m primers may be used in separate compartments to reveal the identity of a chemical entity as well as the point in time of the synthesis of the encoded molecule is has reacted.

In a particular aspect, the invention concerns a method for identifying the chemical entities utilized in the formation of an encoded molecule or a composition of encoded molecules, wherein in separate compartments, n x m primers individually are mixed with an aliquot of a composition obtained by subjecting a library of different complexes to a condition partitioning said composition from the remainder of the library, subjected to a mixture of polymerase and substrate (deoxy)ribonucleotide triphosphates under conditions allowing for an extension reaction to occur when a primer is sufficient complementary to a part of one or more identifier oligonucleotides present in the aliquot, and evaluation, based on measurement of the extension reaction, the presence, absence, or relative abundance of one or more codons in each compartment.

The invention also concerns a set comprising a collection of oligonucleotide primers, a polymerase, a composition of (deoxy)ribonucleotide triphosphates (dNTPs), and a library of complexes composed of a display molecule part and an identifier oligonucleotide, said oligonucleotide comprising codons informative of the identity of the chemical entities which has participated in the formation of the display molecule, wherein the oligonucleotide primers are sufficiently complementary to codons appearing on the identifier oligo nucleotides in the library to allow for an extension to occur.

Detailed Description of the Invention

Complex

The complex comprises an encoded molecule and an identifier oligonucleotide. The identifier comprises codons that identify the encoded molecule. Preferably, the identifier oligonucleotide identifies the encoded molecule uniquely, i.e. in a library of complexes a particular identifier is capable of distinguishing the molecule it is attached to from the rest of the molecules.

The encoded molecule and the identifier may be attached directly to each other or through a bridging moiety. In one aspect of the invention, the bridging moiety is a selectively cleavable linkage.

The identifier oligonucleotide may comprise two or more codons. In a preferred aspect the identifier oligonucleotide comprises three or more codons. The sequence of each codon can be decoded utilizing the present method to identify reactants used in the formation of the encoded molecule. When the identifier comprises more than one codon, each member of a pool of chemical entities can be identified and the order of codons is informative of the synthesis step each member has been incorporated in.

In a certain embodiment, the same codon is used to code for several different chemical entities. In a subsequent identification step, the structure of the encoded molecule can be deduced taking advantage of the knowledge of different attachment chemistries, steric hindrance, deprotection of orthogonal protection groups, etc. In another embodiment, the same codon is used for a group of chemical entities having a common property, such as a lipophilic nature, a certain attachment chemistry etc. In a preferred embodiment, however, the codon is unique i.e. a similar combination of nucleotides does not appear on the identifier oligonucleotide coding for another chemical entity. In a practical approach, for a specific chemical entity, only a single combination of nucleotides is used. In some aspects of the invention, it may be advantageous to use several codons for the same chemical entity, much in the same

way as Nature uses up to six different codons for a single amino acid. The two or more codons identifying the same chemical entity may carry further information related to different reaction conditions.

The sequence of the nucleotides in each codon may have any suitable length. The codon may be a single nucleotide or a plurality of nucleotides. In some aspects of the invention, it is preferred that each codon independently comprises four or more nucleotides, more preferred 4 to 30 nucleotides. In some aspects of the invention the lengths of the codons vary.

A certain codon may be distinguished from any other codon in the library by only a single nucleotide. However, to facilitate a subsequent decoding process and to increase the ability of the primer to discriminate between codons it is in general desired to have two or more mismatches between a particular codon and any other codon appearing on identifier oligonucleotide. As an example, if a codon length of 5 nucleotides is selected, more than 100 nucleotide combinations exist in which two or more mismatches appear. For a certain number of nucleotides in the codon, it is generally desired to optimize the number of mismatches between a particular codon relative to any other codon appearing in the library.

The identifier oligonucleotide will in general have at least two codons arranged in sequence, i.e. next to each other. Two neighbouring codons may be separated by a framing sequence. Depending on the encoded molecule formed, the identifier may comprise further codons, such as 3, 4, 5, or more codons. Each of the further codons may be separated by a suitable framing sequence. Preferably, all or at least a majority of the codons of the identifier are separated from a neighbouring codon by a framing sequence. The framing sequence may have any suitable number of nucleotides, e.g. 1 to 20. Alternatively, codons on the identifier may be designed with overlapping sequences.

The framing sequence, if present, may serve various purposes. In one setup of the invention, the framing sequence identifies the position of the codon. Usually, the framing sequence either upstream or downstream of a codon comprises information which positions the chemical entity and the reaction conditions in the synthesis history of the encodedmolecule. The framing sequence may also or in addition provide for a region of high affinity. The high affinity region may ensure that a hybridisation event with an anti-codon will occur in frame. Moreover, the framing sequence may adjust the annealing temperature to a desired level.

A framing sequence with high affinity can be provided by incorporation of one or more nucleobases forming three hydrogen bonds to a cognate nucleobase. Examples of nucleobases having this property are guanine and cytosine. Alternatively, or in addition, the framing sequence may be subjected to backbone modification. Several back bone modifications provides for higher affinity, such as 2'-O-methyl substitution of the ribose moiety, peptide nucleic acids (PNA), and 2'-4' O-methylene cyclisation of the ribose moiety, also referred to as LNA (Locked Nucleic Acid).

The sequence comprising a codon and an adjacent framing sequence has in a certain aspect of the invention a total length of 11 nucleotides or more, preferably 15 nucleotides or more. A primer may be designed to complementary to the codon sequence as well as the framing sequence. The presence of an extension reaction under conditions allowing for such reaction to occur is indicative of the presence of the chemical entity encoded in the codon as well as the position said chemical entity has in the entire synthesis history of the encoded molecule.

The identifier may comprise flanking regions around the coding section. The flanking regions can also serve as priming sites for amplification reactions, such as PCR or as binding region for oligonucleotide probe. The identifier

may in certain embodiments comprise an affinity region having the property of being able to hybridise to a building block.

It is to be understood that when the term identifier oligonucleotide is used in the present description and claims, the identifier oligonucleotide may be in the sense or the anti-sense format, i.e. the identifier can be a sequence of codons which actually codes for the encoded molecule or can be a sequence complementary thereto. Moreover, the identifier may be single-stranded or double-stranded, as appropriate.

The encoded molecule part of the complex is generally of a structure expected of having an effect on a target. When the target is of pharmaceutical importance, the encoded molecule is generally a possible drug candidate. The complex may be formed by tagging a library of different possible drug candidates with a tag, e.g. a nucleic acid tag identifying each possible drug candidate. In another embodiment of the invention, the molecule formed by a variety of reactants which have reacted with each other and/or a scaffold molecule. Optionally, this reaction product may be post-modified to obtain the final molecule displayed on the complex. The post-modification may involve the cleavage of one or more chemical bonds attaching the encoded molecule to the identifier in order more efficiently to display the encoded molecule.

The formation of an encoded molecule generally starts by a scaffold, i.e. a chemical unit having one or more reactive groups capable of forming a connection to another reactive group positioned on a chemical entity, thereby generating an addition to the original scaffold. A second chemical entity may react with a reactive group also appearing on the original scaffold or a reactive group incorporated by the first chemical entity. Further chemical entities may be involved in the formation of the final reaction product. The formation of a connection between the chemical entity and the nascent encoded molecule may be mediated by a bridging molecule. As an example, if the nascent encoded molecule and the chemical entity both comprise an

amine group a connection between these can be mediated by a dicarboxylic acid. A synthetic molecule is in general produced in vitro and may be a naturally occurring or an artificial substance. Usually, a synthetic molecule is not produced using the naturally translation system in an in vitro process.

The chemical entities that are precursors for structural additions or eliminations of the encoded molecule may be attached to a building block prior to the participation in the formation of the reaction product leading the final encoded molecule. Besides the chemical entity, the building block generally comprises an anti-codon. In some embodiments the building blocks also comprise an affinity region providing for affinity towards the nascent complex.

Thus, the chemical entities are suitably mediated to the nascent encoded molecule by a building block, which further comprises an anticodon. The anti-codon serves the function of transferring the genetic information of the building block in conjunction with the transfer of a chemical entity. The transfer of genetic information and chemical entity may occur in any order. The chemical entities are preferably reacted without enzymatic interaction in some aspects of the invention. Notably, the reaction of the chemical entities is preferably not mediated by ribosomes or enzymes having similar activity. In other aspects of the invention, enzymes are used to mediate the reaction between a chemical entity and a nascent encoded molecule.

According to certain aspects of the invention the genetic information of the anti-codon is transferred by specific hybridisation to a codon on a nucleic acid template. Another method for transferring the genetic information of the anti-codon to the nascent complex is to anneal an oligonucleotide complementary to the anti-codon and attach this oligonucleotide to the complex, e.g. by ligation. A still further method involves transferring the genetic information of the anti-codon to the nascent complex by an extension reaction using a polymerase and a mixture of dNTPs.

The chemical entity of the building block may in most cases be regarded as a precursor for the structural entity eventually incorporated into the encoded molecule. In other cases the chemical entity provides for the eliminations of chemical units of the nascent encoded molecule. Therefore, when it in the present application with claims is stated that a chemical entity is transferred to a nascent encoded molecule it is to be understood that not necessarily all the atoms of the original chemical entity is to be found in the eventually formed encoded molecule. Also, as a consequence of the reactions involved in the connection, the structure of the chemical entity can be changed when it appears on the nascent encoded molecule. Especially, the cleavage resulting in the release of the entity may generate a reactive group which in a subsequent step can participate in the formation of a connection between a nascent complex and a chemical entity.

The chemical entity of the building block comprises at least one reactive group capable of participating in a reaction which results in a connection between the chemical entity of the building block and another chemical entity or a scaffold associated with the nascent complex. The number of reactive groups which appear on the chemical entity is suitably one to ten. A building block featuring only one reactive group is used i.a. in the end positions of polymers or scaffolds, whereas building blocks having two reactive groups are suitable for the formation of the body part of a polymer or scaffolds capable of being reacted further. One, two or more reactive groups intended for the formation of connections, are typically present on scaffolds. Non-limiting examples of scaffolds are opiates, steroids, benzodiazepines, hydantoines, and peptidylphosphonates.

The reactive group of the chemical entity may be capable of forming a direct connection to a reactive group of the nascent complex or the reactive group of the building block may be capable of forming a connection to a reactive group of the nascent complex through a bridging fill-in group. It is to be un-

derstood that not all the atoms of a reactive group are necessarily maintained in the connection formed. Rather, the reactive groups are to be regarded as precursors for the structure of the connection.

The subsequent cleavage step to release the chemical entity from the building block can be performed in any appropriate way. In an aspect of the invention the cleavage involves usage of a chemical reagent or an enzyme. The cleavage results in a transfer of the chemical entity to the nascent encoded molecule or in a transfer of the nascent encoded molecule to the chemical entity of the building block. In some cases it may be advantageous to introduce new chemical groups as a consequence of linker cleavage. The new chemical groups may be used for further reaction in a subsequent cycle, either directly or after having been activated. In other cases it is desirable that no trace of the linker remains after the cleavage.

In another aspect, the connection and the cleavage is conducted as a simultaneous reaction, i.e. either the chemical entity of the building block or the nascent encoded molecule is a leaving group of the reaction. In some aspects of the invention, it is appropriate to design the system such that the connection and the cleavage occur simultaneously because this will reduce the number of steps and the complexity. The simultaneous connection and cleavage can also be designed such that either no trace of the linker remains or such that a new chemical group for further reaction is introduced, as described above.

The attachment of the chemical entity to the building block, optionally via a suitable spacer can be at any entity available for attachment, e.g. the chemical entity can be attached to a nucleobase or the backbone. In general, it is preferred to attach the chemical entity at the phosphor of the internucleoside linkage or at the nucleobase. When the nucleobase is used for attachment of the chemical entity, the attachment point is usually at the 7 position of the purines or 7-deaza-purins or at the 5 position of pyrimidines. The nucleotide

may be distanced from the reactive group of the chemical entity by a spacer moiety. The spacer may be designed such that the conformational spaced sampled by the reactive group is optimized for a reaction with the reactive group of the nascent encoded molecule.

The encoded molecules may have any chemical structure. In a preferred aspect, the encoded molecule can be any compound that may be synthesized in a component-by-component fashion. In some aspects the synthetic molecule is a linear or branched polymer. In another aspect the synthetic molecule is a scaffolded molecule. The term "encoded molecule" also comprises naturally occurring molecules like α-polypeptides etc, however produced *in vitro* usually in the absence of enzymes, like ribosomes. In certain aspects, the synthetic molecule of the library is a non-α-polypeptide.

The encoded molecule may have any molecular weight. However, in order to be orally available, it is in this case preferred that the synthetic molecule has a molecular weight less than 2000 Daltons, preferably less than 1000 Dalton, and more preferred less than 500 Daltons.

The size of the library may vary considerably pending on the expected result of the inventive method. In some aspects, it may be sufficient that the library comprises two, three, or four different complexes. However, in most events, more than two different complexes are desired to obtain a higher diversity. In some aspects, the library comprises 1,000 or more different complexes, more preferred 1,000,000 or more different complexes. The upper limit for the size of the library is only restricted by the size of the vessel in which the library is comprised. It may be calculated that a vial may comprise up to 10^{14} different complexes.

Extension reaction

The extension reaction requires a primer, a polymerase as well as a collection of deoxyribonucleotide triphosphates (abbreviated dNTP's herein) to proceed. An extension product may be obtained in the event the primer is sufficient complementary to an identifier oligonucleotide for a polymerase to recognise the double helix as a substrate. After binding of the polymerase to the double helix, the deoxyribonucleotide triphosphates (blend of dATP, dCTP, dGTP, and dTTP) are incorporated into the extension product using the identifier oligonucleotide as template. The conditions allowing for the extension reaction to occur usually includes a suitable buffer. The buffer may be any aqueous or organic solvent or mixture of solvents in which the polymerase has a sufficient activity. To facilitate the extension process the polymerase and the mixture of dNTP's are generally included in a buffer which is added to the identifier oligonucleotide and primer mixture. An exemplary kit comprising the polymerase and the nNTP's for performing the extension process comprises the following: 50 mM KCl; 10 mM Tris-HCl at pH 8.3; 1.5 mM MgCl2 ; 0.001% (wt/vol) gelatin, 200 μM dATP; 200 μM dTTP; 200 μM dCTP; 200 μM dGTP; and 2.5 units Thermus aquaticus (Taq) DNA polymerase I (U.S. Pat. No. 4,889,818) per 100 microliters (μl) of buffer.

The primer may be selected to be complementary to one or more codons or parts of such codons. The length of the primers may be determined by the length of the codons, however, the primers usually are at least about 11 nucleotides in length, more preferred at least 15 nucleotides in length to allow for an efficient extension by the polymerase. The presence or absence of one or more codons is indicated by the presence of or absence of an extension product. The extension product may be measured by any suitable method, such as size fractioning on an agarose gel and staining with ethidium bromide.

In a preferred embodiment the admixture of identifier oligonucleotide and primer is termocycled to obtain a sufficient number of copies of the extension

product. The thermocycling is typically carried out by repeatedly increasing and decreasing the temperature of the mixture within a temperature range whose lower limit is about 30 degrees Celsius (30°C) to about 55°C and whose upper limit is about 90°C to about 100° C. The increasing and decreasing can be continuous, but is preferably phasic with time periods of relative temperature stability at each of temperatures favouring polynucleotide synthesis, denaturation and hybridization.

When a single complex is analysed in accordance with the present method, the result may be used to verify the presence or absence of a specific chemical entity during the formation of the display molecule. The formation of an extension product is indicative of the presence of an oligonucleotide part complementary to the primer in the identifier oligonucleotide. Conversely, the absence of an extension product is indicative of the absence of an oligonucleotide part complementary to the primer in the identifier oligonucleotide. Selecting the sequence of the primer such that it is complementary to one or more codons will therefore provide information of the structure of the encoded molecule coded for by this codon(s).

In a preferred aspect of the invention, in the mixture of the identifier oligonucleotide and the primer oligonucleotide, a second primer complementary to a sequence of the extension product is included. The second primer is also termed reverse primer and ensures an exponential increase of the number of produced extension products. The method using a forward and reverse primer is well known to skilled person in the art and is generally referred to as polymerase chain reaction (abbreviated PCR) in the present application with claims. In one embodiment of the invention the reverse primer is annealed to a part of the extension product downstream, i.e. near the 3'end of the extension product, or a part complementing the coding part of the identifier oligonucleotide. In another embodiment, the first primer (forward primer) anneals to an upstream position of the identifier oligonucleotide, preferably be-

fore the coding part, and the reverse primer anneals to a sequence of the extension product complementing one or more codons or parts thereof.

The amplicons resulting from the PCR process may be stained during or following the reaction to ease the detection. A staining after the PCR process may be prepared with e.g. ethidium bromide or a similar staining agent. As an example, amplicons from the PCR process is run on an agarose gel and subsequently stained with ethidium bromide. Under UV illumination bands of amplicons becomes visible. It is possible to incorporate the staining agent in the agarose gel or to allow a solution of the staining agent to migrate through the gel. The amplicons may also be stained during the PCR process by an intercalating agent, like CYBR. In presence of the intercalating agent while the amplification proceeds it will incorporate in the double helix. The intercalation agent may then be made visible by irradiation by a suitable source.

The intensity of the staining is informative of the relative abundance of a specific amplicon. Thus, it is possible to quantify the occurrence of a codon in an identifier oligonucleotide. When a library of bifunctional complexes has been subjected to a selection the codons in the pool of identifier oligonucleotides which has been selected can be quantified using this method. As an example a sample of the selected identifier oligonucleotides is subjected to various PCR amplifications with different primers in separate compartments and the PCR product of each compartment is analysed by electrophoresis in the presence of ethidium bromide. The bands that appear can be quantified by a densitometric analysis after irradiation by ultraviolet light and the relative abundance of the codons can be measured.

Alternatively, the primers may be labelled with a suitable small molecule, like biotin or digoxigenin. A PCR-ELISA analysis may subsequently be performed based on the amplicons comprising the small molecule. A preferred method includes the application of a solid support covered with streptavidin or avidin when biotin is used as label and anti-digoxigenin when digoxigenin is used as

the label. Once captured, the amplicons can be detected using an enzyme-labelled avidin or anti-dixigenin reporter molecule similar to a standard ELISA format.

To avoid laborious post-PCR handling steps required to evaluate the amplicons, it is in a certain embodiment preferred to measure the extension process "real time". Several real time PCR processes has been developed and all the suitable real time PCR process available to the skilled person in the art can be used in the evaluating step of the present invention and are include in the present scope of protection. The PCR reactions discussed below are of particular interest.

The monitoring of accumulating amplicons in real time has been made possible by labelling of primers, probes, or amplicons with fluorogenic molecules. The real time PCR amplification is usually performed with a speed faster than the conventional PCR, mainly due to reduced cycles time and the use of sensitive methods for detection of emissions from the fluorogenic labels. The most commonly used fluorogenic oligoprobes rely upon fluorescent resonance energy transfer (FRET) between fluorogenic labels or between one flourophor and a dark or "black-hole" non-fluorescent quencher (NFQ), which disperse energy as heat rather than fluorescence. FRET is a spectroscopic process by which energy is passed between molecules separated by 10-100 Å that have overlapping emission and absorption spectra. An advantage of many real time PCR methods is that they can be carried out in a closed system, i.e. a system which does not need to be opened to examine the result of the PCR. A closed system implies a reduced result turnaround, minimisation of the potential for carry-over contamination and the ability to closely scrutinise the essay's performance.

The present real time PCR methods currently available to the skilled person can be classified into either amplicon sequence specific or non-specific methods. The basis for the non-specific detection methods is a DNA-binding

fluorogenic molecule. Included in this class are the earliest and simplest approaches to real time PCR. Ethidium bromide, YO-PRO-1, and SYBR® green 1 all fluorescence when associated with double stranded DNA which is exposed to a suitable wavelength of light. This approach requires the fluorescent agent to be present during the PCR process and provides for a real time detection of the fluorescent agent as it is incorporated into the double stranded helix.

The amplicons sequence specific methods includes, but are not limited to, the TaqMan®, hairpin, LightCycler®, Sunrise®, and Scorpion® methods. The LightCycler® method also designated "HybProbes" make use of a pair of adjacent, fluorogenic hybridisation oligonucleotide probes. A first, usually the upstream oligoprobe is labelled with a 3' donor fluorophore and the second, usually the downstream probe is commonly labelled with either a Light cycler Red 640 or Red 705 acceptor fluorophore a the 5' terminus so that when both oligoprobes are hybridised the two fluorophores are located in close proximity, such as within 10 nm, of each other. The close proximity provides for the emission of a fluorescence when irradiated with a suitable light source, such a blue diode in case of the LightCycler®. The region for annealing of the probes may be any suitable position that does not interfere with the primer annealing. In a suitable setup, the site for binding the probes are positioned downstream of the codon region on the identifier oligonucleotide. Alternatively, when a reverse primer is used, the region for annealing the probes may be at the 3' end of the strand complementing the identifier oligonucleotide. Another embodiment of the LightCycler method includes that the pair of oligonucleotide probes are annealed to one or more codons and primer sites exterior to the coding part of the identifier oligonucleotide are used for PCR amplification.

The TaqMan® method, also referred to as the 5' nuclease or hydrolysis method, requires an oligoprobe, which is attached to a reporter flourophor, such as 6-carboxy-fluoroscein, and a quencher fluorophore, such as 6-

carboxy-tetramethyl-rhodamine, at each end. When in close proximity, i.e. annealed to an identifier oligonucleotide, or a sequence complementing the identifier oligonucleotide, the quencher will "hijack" the emissions that have resulted from the excitation of the reporter. As the polymerase progresses along the relevant strand, it displaces and the hydrolyses the oligoprobe via its 5'→3' endonuclease activity. Once the reporter is removed from the extinguishing influence of the quencher, it is able to release excitation energy at a wavelength that can be monitored by a suitable instrument, such as ABI Prism® 7700. The fractional cycle number at which the real-time fluorescence signal mirrors progression of the reaction above the background noise is normally used as an indicator of successful identifier oligonucleotide amplification. This threshold cycle (C_T) is defined as the PCR cycle in which the gain in fluorescence generated by the accumulating amplicons exceeds 10 standard deviations of the mean base line fluorescence. The C_T is proportional to the number of identifier oligonucleotide copies present in the sample. The TaqMan probe is usually designed to hybridise at a position downstream of a primer binding site, be it a forward or a reverse primer. When the primer is designed to anneal to one or more codons of the identifier oligonucleotide, the presence of these one or more codons is indicated by the emittance of light. Furthermore, the quantity of the identifier oligonucleotides comprising the one or more codons may be measured by the C_T value.

The Hairpin method involves an oligoprobe, in which a fluorophore and a quencher are positioned at the termini. The labels are hold in close proximity by distal stem regions of homologous base pairing deliberately designed to create a hairpin structure which result in quenching either by FRET or a direct energy transfer by a collisional mechanism due to the intimate proximity of the labels. When direct energy transfer by a collision mechanism is used the quencher is usually different from the FRET mechanism, and is suitably 4-(4'-dimethylamino-phenylazo)-benzene (DABCYL). In the presence of a complementary sequence, usually downstream of a primer, or within the bounds of the primer binding sides in case of more than one a single primer,

the oligoprobe will hybridise, shifting into an open configuration. The fluorophore is now spatially removed from the quencher's influence and fluorescence emissions are monitored during each cycle. In a certain aspect, the hairpin probe may be designed to anneal to a codon in order to detect this codon if present on the identifier oligonucleotide. This embodiment may be suitable if codons only differs from each other with a single or a few nucleotides, because is in well-known that the occurrence of a mismatch between a hairpin oligoprobe and its target sequence has a greater destabilising effect on the duplex than the introduction of an equivalent mismatch between the target oligonucleotide and a linear oligoprobe. This is probably because the hairpin structure provides a highly stable alternate conformation.

The Sunrise and Scorpion methods are similar in concept to the hairpin oligoprobe, except that the label becomes irreversible incorporated in to the PCR product. The Sunrise method involves a primer (commercially available as Amplifluor™ hairpin primers) comprising a 5' fluorophore and a quencher, e.g. DABCYL. The labels are separated by complementary stretches of sequence that create a stem when the sunrise primer is closed. At the 3' terminus is a target specific primer sequence. In a preferred embodiment the target sequence is a codon, optionally more codons. The sunrise primer's sequence is intended to be duplicated by the nascent complementary stand and, in this way, the stem is destabilised, the two fluorophores are held apart, usually between 15 and 25 nucleotides, and the fluorophore is free to emit its excitation energy for monitoring. The Scorpion primer resembles the sunrise primer, but derivate in having a moiety that blocks duplication on the signalling portion of the scorpion primer. The blocking moiety is typically hexethylene glycol. In addition to the difference in structure, the function of the scorpion primers differs slightly in that the 5' region of the oligonucleotide is designed to hybridise to a complementary region within the amplicons. In a certain embodiment the complementary region is a codon on the identifier oligonucleotide. The hybridisation forces the labels apart disrupting the hairpin and permitting emission in the same way as the hairpin probes.

Methods for forming a library of complexes

The complexes comprising an identifier part having two or more codons that codes for reactants that have reacted in the formation of the encoded molecule part of the complex may be formed by a variety of processes. Generally, the preferred methods can be used for the formation of virtually any kind of encode molecule. Suitable examples of processes include prior art methods disclosed in WO 93/20242, WO 93/06121, WO 00/23458, WO 02/074929, and WO 02/103008, the content of which being incorporated herein by reference as well as methods of the present applicant not yet public available, including the methods disclosed in DK PA 2002 01955 filed 19 December 2002, and DK PA 2003 00430 filed 20 March 2003. Any of these methods may be used, and the entire content of the patent applications are included herein by reference.

Below four preferred embodiments are described. A first embodiment disclosed in more detail in WO 02/103008 is based on the use of a polymerase to incorporate unnatural nucleotides as building blocks. Initially, a plurality of template oligonucleotides is provided. Subsequently primers are annealed to each of the templates and a polymerase is extending the primer using nucleotide derivatives which have appended chemical entities. Subsequent to or simultaneously with the incorporation of the nucleotide derivatives, the chemical entities are reacted to form a reaction product. The encoded molecule may be post-modified by cleaving some of the linking moieties to better present the encoded molecule.

Several possible reaction approaches for the chemical entities are apparent. First, the nucleotide derivatives can be incorporated and the chemical entities subsequently polymerised. In the event the chemical entities each carry two reactive groups, the chemical entities can be attached to adjacent chemical entities by a reaction of these reactive groups. Exemplary of the reactive groups are amine and carboxylic acid, which upon reaction form an amide

bond. Adjacent chemical entities can also be linked together using a linking or bridging moiety. Exemplary of this approach is the linking of two chemical entities each bearing an amine group by a bi-carboxylic acid. Yet another approach is the use of a reactive group between a chemical entity and the nucleotide building block, such as an ester or a thioester group. An adjacent building block having a reactive group such as an amine may cleave the interspaced reactive group to obtain a linkage to the chemical entity, e.g. by an amide linking group.

A second embodiment for obtainment of complexes pertains to the use of hybridisation of building blocks to an identifier oligonucleotide and reaction of chemical entities attached to the building blocks in order to obtain a reaction product. This approach comprises that templates are contacted with a plurality of building blocks, wherein each building block comprises an anti-codon and a chemical entity. The anti-codons are designed such that they recognise a sequence, i.e. a codon, on the template. Subsequent to the annealing of the anti-codon and the codon to each other a reaction of the chemical entity is effected.

The template may be associated with a scaffold. Building blocks bringing chemical entities in may be added sequentially or simultaneously and a reaction of the reactive group of the chemical entity may be effected at any time after the annealing of the building blocks to the template.

A third embodiment for the generation of a complex includes chemical or enzymatical ligation of building blocks when these are lined up on a template. Initially, templates are provided, each having one or more codons. The templates are contacted with building blocks comprising anti-codons linked to chemical entities. The two or more anti-codons annealed on a template are subsequently ligated to each other and a reaction of the chemical entities is effected to obtain a reaction product. The method is disclosed in more detail in DK PA 2003 00430 filed 20 March 2003.

A fourth embodiment makes use of the extension by a polymerase of an affinity sequence of the nascent complex to transfer the anti-codon of a building block to the nascent complex. The method implies that a nascent complex comprising a scaffold and an affinity region is annealed to a building block comprising a region complementary to the affinity section. Subsequently the anti-codon region of the building block is transferred to the nascent complex by a polymerase. The transfer of the chemical entity may be transferred prior to, simultaneously with or subsequent to the transfer of the anti-codon. This method is disclosed in detail in DK PA 2002 01955 filed 19 December 2002.

After or simultaneously with the formation of the reaction product some of the linkers to the template may be cleaved, however at least one linker must be maintained to provide for the complex.

Nucleotides

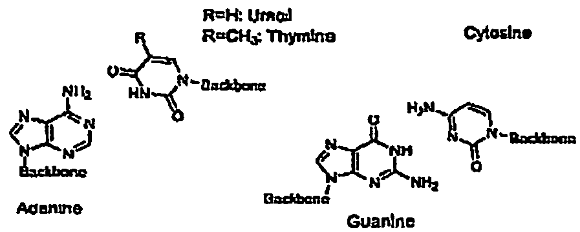
The nucleotides used in the present invention may be linked together in a sequence of nucleotides, i.e. an oligonucleotide. Each nucleotide monomer is normally composed of two parts, namely a nucleobase moiety, and a back-bone. The back bone may in some cases be subdivided into a sugar moiety and an internucleoside linker.

The nucleobase moiety may be selected among naturally occurring nucleo-bases as well as non-naturally occurring nucleobases. Thus, "nucleobase" includes not only the known purine and pyrimidine hetero-cycles, but also heterocyclic analogues and tautomers thereof. Illustrative examples of nu-cleobases are adenine, guanine, thymine, cytosine, uracil, purine, xanthine, diaminopurine, 8-oxo-N^6-methyladenine, 7-deazaxanthine, 7-deazaguanine, N^4,N^4-ethanocytosin, N^6,N^6-ethano-2,6-diamino-purine, 5-methylcytosine, 5-(C^3-C^6)-alkynylcytosine, 5-fluorouracil, 5-bromouracil, pseudoisocytosine, 2-hydroxy-5-methyl-4-triazolopyridine, isocytosine, isoguanine, inosine and the

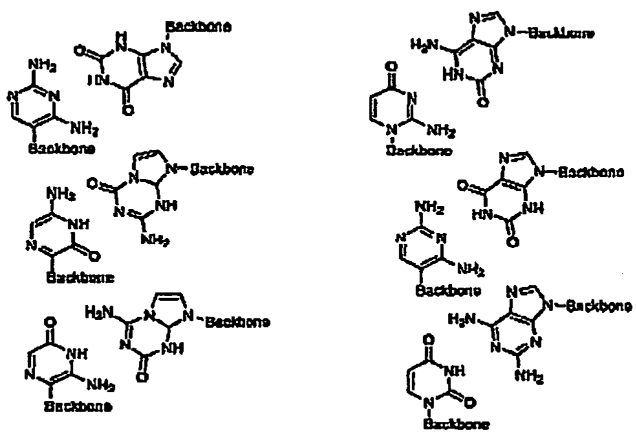
"non-naturally occurring" nucleobases described in Benner et al., U.S. Pat No. 5,432,272. The term "nucleobase" is intended to cover these examples as well as analogues and tautomers thereof. Especially interesting nucleobases are adenine, guanine, thymine, cytosine, 5-methylcytosine, and uracil, which are considered as the naturally occurring nucleobases in relation to therapeutic and diagnostic application in humans.

Examples of suitable specific pairs of nucleobases are shown below:

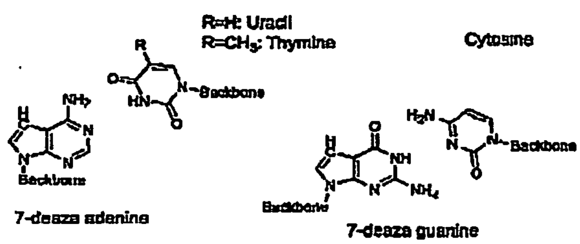
Natural Base Pairs



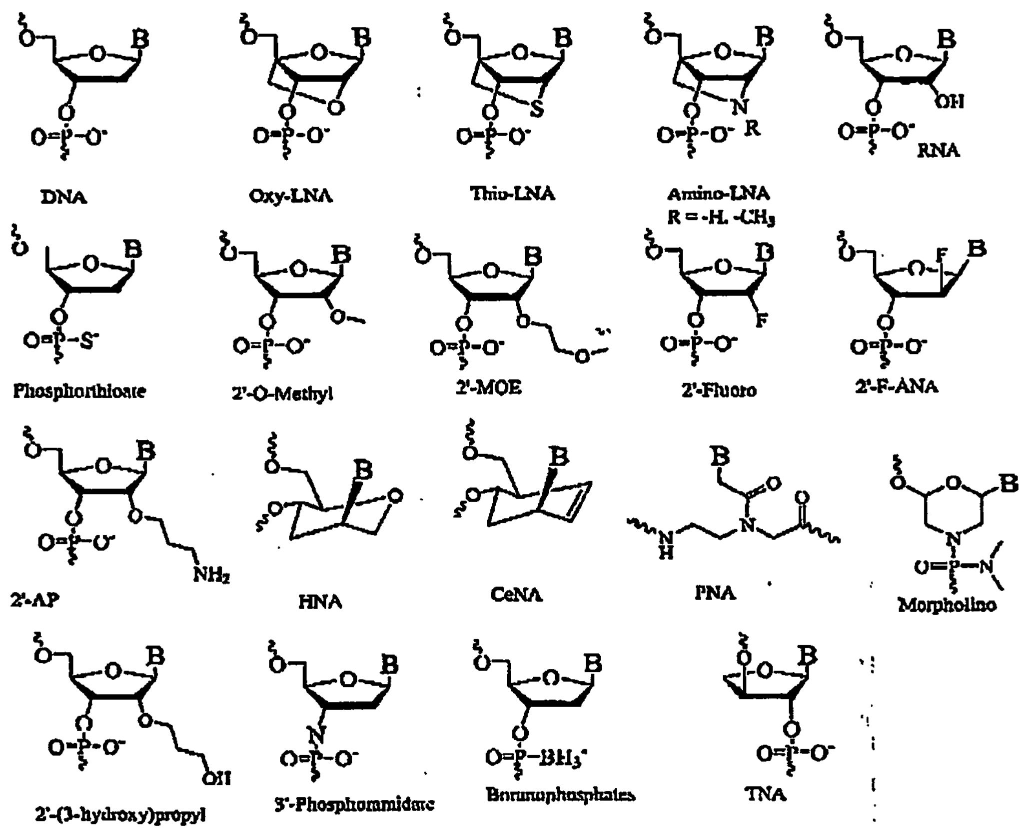
Synthetic Base Pairs



Synthetic purine bases pairing with natural pyrimidines



Suitable examples of backbone units are shown below (B denotes a nucleobase):



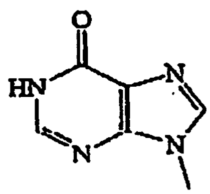
The sugar moiety of the backbone is suitably a pentose but may be the appropriate part of an PNA or a six-member ring. Suitable examples of possible pentoses include ribose, 2'-deoxyribose, 2'-O-methyl-ribose, 2'-flour-ribose, and 2'-4'-O-methylene-ribose (LNA). Suitably the nucleobase is attached to the 1' position of the pentose entity.

An internucleoside linker connects the 3' end of preceding monomer to a 5' end of a succeeding monomer when the sugar moiety of the backbone is a pentose, like ribose or 2-deoxyribose. The internucleoside linkage may be the natural occurring phospodiester linkage or a derivative thereof. Examples of such derivatives include phosphorothioate, methylphosphonate, phosphoramidate, phosphotriester, and phosphodithioate. Furthermore, the

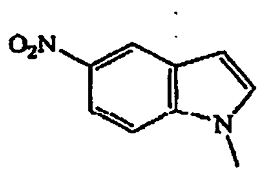
internucleoside linker can be any of a number of non-phosphorous-containing linkers known in the art.

Preferred nucleic acid monomers include naturally occurring nucleosides forming part of the DNA as well as the RNA family connected through phosphodiester linkages. The members of the DNA family include deoxyadenosine, deoxyguanosine, deoxythymidine, and deoxycytidine. The members of the RNA family include adenosine, guanosine, uridine, cytidine, and inosine. Inosine is a non-specific pairing nucleoside and may be used as universal base because inosine can pair nearly isoenergetically with A, T, and C. Other compounds having the same ability of non-specifically base-pairing with natural nucleobases have been formed. Suitable compounds which may be utilized in the present invention includes among others the compounds depicted below

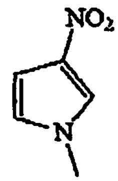
Examples of Universal Bases:



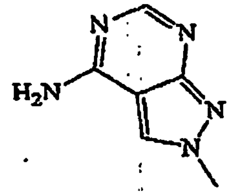
Inosine



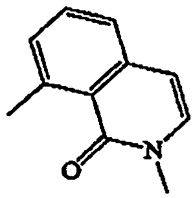
5-Nitroindole



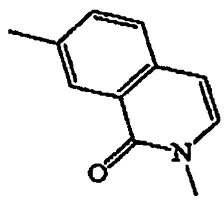
3-Nitropyrrole



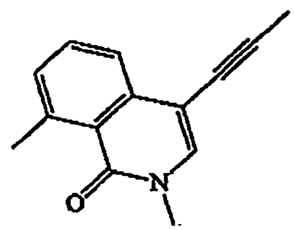
N^8-8aza-7deazaadenine



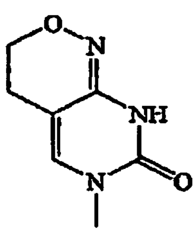
MICS



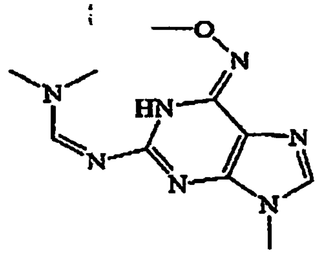
5MICS



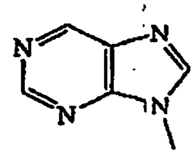
PIM



dP



dK



Nebularine

Building block

The chemical entities that are precursors for structural additions or eliminations of the encoded molecule may be attached to a building block prior to

the participation in the formation of the reaction product leading the final encoded molecule. Besides the chemical entity, the building block generally comprises an anti-codon.

The chemical entity of the building block comprises at least one reactive group capable of participating in a reaction which results in a connection between the chemical entity of the building block and another chemical entity or a scaffold associated with the nascent complex. The connection is facilitated by one or more reactive groups of the chemical entity. The number of reactive groups which appear on the chemical entity is suitably one to ten. A building block featuring only one reactive group is used i.a. in the end positions of polymers or scaffolds, whereas building blocks having two reactive groups are suitable for the formation of the body part of a polymer or scaffolds capable of being reacted further. One, two or more reactive groups intended for the formation of connections, are typically present on scaffolds.

The reactive group of the building block may be capable of forming a direct connection to a reactive group of the nascent complex or the reactive group of the building block may be capable of forming a connection to a reactive group of the nascent complex through a bridging fill-in group. It is to be understood that not all the atoms of a reactive group are necessarily maintained in the connection formed. Rather, the reactive groups are to be regarded as precursors for the structure of the connection.

The subsequent cleavage step to release the chemical entity from the building block can be performed in any appropriate way. In an aspect of the invention the cleavage involves usage of a reagent or an enzyme. The cleavage results in a transfer of the chemical entity to the nascent encoded molecule or in a transfer of the nascent encoded molecule to the chemical entity of the building block. In some cases it may be advantageous to introduce new chemical groups as a consequence of linker cleavage. The new chemical groups may be used for further reaction in a subsequent cycle, either directly

or after having been activated. In other cases it is desirable that no trace of the linker remains after the cleavage.

In another aspect, the connection and the cleavage is conducted as a simultaneous reaction, i.e. either the chemical entity of the building block or the nascent encoded molecule is a leaving group of the reaction. In general, it is preferred to design the system such that the connection and the cleavage occur simultaneously because this will reduce the number of steps and the complexity. The simultaneous connection and cleavage can also be designed such that either no trace of the linker remains or such that a new chemical group for further reaction is introduced, as described above.

The attachment of the chemical entity to the building block, optionally via a suitable spacer can be at any entity available for attachment, e.g. the chemical entity can be attached to a nucleobase or the backbone. In general, it is preferred to attach the chemical entity at the phosphor of the internucleoside linkage or at the nucleobase. When the nucleobase is used for attachment of the chemical entity, the attachment point is usually at the 7 position of the purines or 7-deaza-purins or at the 5 position of pyrimidines. The nucleotide may be distanced from the reactive group of the chemical entity by a spacer moiety. The spacer may be designed such that the conformational space sampled by the reactive group is optimized for a reaction with the reactive group of the nascent encoded molecule or reactive site.

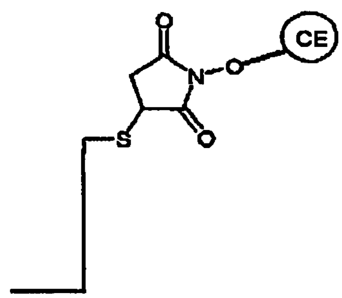
The anticodon complements the codon of the identifier sequence and generally comprises the same number of nucleotides as the codon. The anticodon may be adjoined with a fixed sequence, such as a sequence complementing a framing sequence.

Various specific building blocks are envisaged. Building blocks of particular interest are shown below.

Specific Building blocks

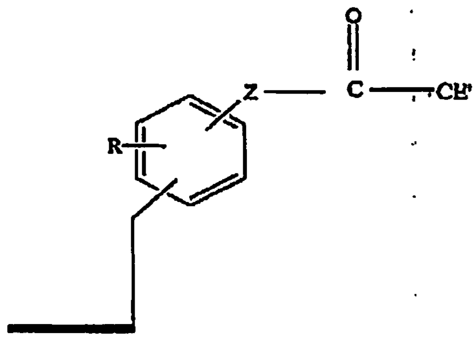
Building blocks transferring a chemical entity to a recipient nucleophilic group

The building block indicated below is capable of transferring a chemical entity (CE) to a recipient nucleophilic group, typically an amine group. The bold lower horizontal line illustrates the building block and the vertical line illustrates a spacer. The 5-membered substituted N-hydroxysuccinimid (NHS) ring serves as an activator, i.e. a labile bond is formed between the oxygen atom connected to the NHS ring and the chemical entity. The labile bond may be cleaved by a nucleophilic group, e.g. positioned on a scaffold



The 5-membered substituted N-hydroxysuccinimid (NHS) ring serves as an activator, i.e. a labile bond is formed between the oxygen atom connected to the NHS ring and the chemical entity. The labile bond may be cleaved by a nucleophilic group, e.g. positioned on a scaffold, to transfer the chemical entity to the scaffold, thus converting the remainder of the fragment into a leaving group of the reaction. When the chemical entity is connected to the activator through an carbonyl group and the recipient group is an amine, the bond formed on the scaffold will an amide bond. The above building block is the subject of the Danish patent application No. PA 2002 01946 and the US provisional patent application No. 60/434,439, the content of which are incorporated herein in their entirety by reference.

Another building block which may form an amide bond is

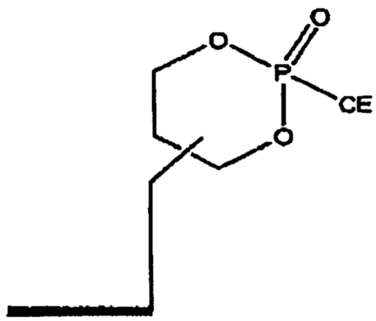


R may be absent or NO_2, CF_3, halogen, preferably Cl, Br, or I, and Z may be S or O. This type of building block is disclosed in Danish patent application No. PA 2002 0951 and US provisional patent application filed 20 December 2002 with the title "A building block capable of transferring a functional entity to a recipient reactive group". The content of both patent application are incorporated herein in their entirety by reference.

A nucleophilic group can cleave the linkage between Z and the carbonyl group thereby transferring the chemical entity –(C=O)-CE' to said nucleophilic group.

Building blocks transferring a chemical entity to a recipient reactive group forming a C=C bond

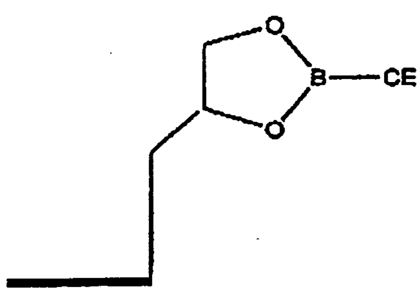
A building block as shown below are able to transfer the chemical entity to a recipient aldehyde group thereby forming a double bond between the carbon of the aldehyde and the chemical entity



The above building block is comprised by the Danish patent application No. DK PA 2002 01952 and the US provisional patent application filed 20 December 2002 with the title "A building block capable of transferring a functional entity to a recipient reactive group forming a C=C double bond". The content of both patent applications are incorporated herein in their entirety by reference.

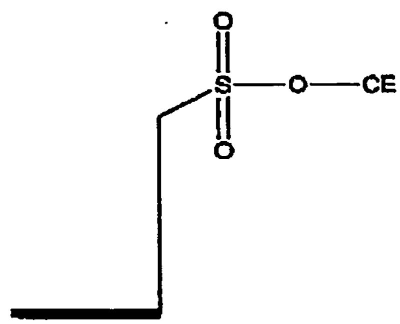
Building blocks transferring a chemical entity to a recipient reactive group forming a C-C bond

The below building block is able to transfer the chemical entity to a recipient group thereby forming a single bond between the receiving moiety, e.g. a scaffold, and the chemical entity.



The above building block is comprised by the Danish patent application No. DK PA 2002 01947 and the US provisional patent application No 60/434,428. The content of both patent applications are incorporated herein in their entirety by reference.

Another building block capable of transferring a chemical entity to a receiving reactive group forming a single bond is



The receiving group may be a nucleophile, such as a group comprising a hetero atom, thereby forming a single bond between the chemical entity and the hetero atom, or the receiving group may be an electronegative carbon atom, thereby forming a C-C bond between the chemical entity and the scaffold.

The chemical entity attached to any of the above building blocks may be a selected from a large arsenal of chemical structures. Examples of chemical entities are
H or entities selected among the group consisting of a C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_4-C_8 alkadienyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R^4, 0-3 R^5 and 0-3 R^9 or C_1-C_3 alkylene-NR^4_2, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 alkylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9.

where R^4 is H or selected independently among the group consisting of C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, heteroaryl, said group being substituted with 0-3 R^9 and

R^5 is selected independently from $-N_3$, -CNO, $-C(NOH)NH_2$, -NHOH, $-NHNHR^6$, $-C(O)R^6$, $-SnR^6_3$, $-B(OR^6)_2$, $-P(O)(OR^6)_2$ or the group consisting of

C_2-C_8 alkenyl, C_2-C_6 alkynyl, C_4-C_8 alkadienyl said group being substituted with 0-2 R^7,

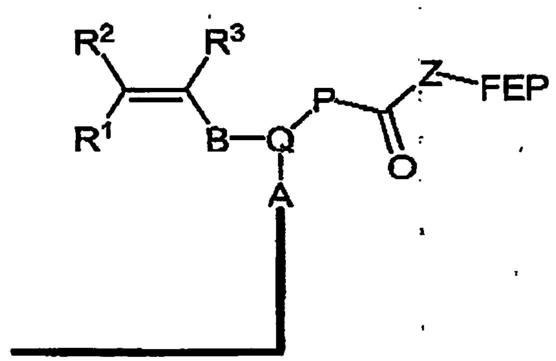
where R^6 is selected independently from H, C_1-C_6 alkyl, C_3-C_7 cycloalkyl, aryl or C_1-C_6 alkylene-aryl substituted with 0-5 halogen atoms selected from -F, -Cl, -Br, and –I; and R^7 is independently selected from $-NO_2$, $-COOR^6$, $-COR^6$, -CN, $-OSiR^6_3$, $-OR^6$ and $-NR^6_2$.

R^8 is H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_7 cycloalkyl, aryl or C_1-C_6 alkylene-aryl substituted with 0-3 substituents independently selected from –F, -Cl, $-NO_2$, $-R^3$, $-OR^3$, $-SiR^3_3$

R^9 is =O, -F, -Cl, -Br, -I, -CN, $-NO_2$, $-OR^6$, $-NR^6_2$, $-NR^6$-C(O)R^8, $-NR^6$-C(O)OR^8, $-SR^6$, -S(O)R^6, $-S(O)_2R^6$, $-COOR^6$, -C(O)NR^6_2 and $-S(O)_2NR^6_2$.

Cross-link cleavage building blocks

It may be advantageous to split the transfer of a chemical entity to a recipient reactive group into two separate steps, namely a cross-linking step and a cleavage step because each step can be optimized. A suitable building block for this two step process is illustrated below:



Initally, a reactive group appearing on the functional entity precursor (abbreviated FEP) reacts with a recipient reactive group, e.g. a reactive group appearing on a scaffold, thereby forming a cross-link. Subsequently, a cleavage is performed, usually by adding an aqueous oxidising agent such as I_2, Br_2,

Cl_2, H^+, or a Lewis acid. The cleavage results in a transfer of the group HZ-FEP- to the recipient moiety, such as a scaffold.

In the above formula

Z is O, S, NR^4

Q is N, CR^1

P is a valence bond, O, S, NR^4, or a group C_{5-7}arylene, C_{1-6}alkylene, C_{1-6}O-alkylene, C_{1-6}S-alkylene, NR^1-alkylene, C_{1-6}alkylene-O, C_{1-6}alkylene-S option said group being substituted with 0-3 R^4, 0-3 R^5 and 0-3 R^9 or C_1-C_3 alkylene-NR^4_2, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 alkylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9,

B is a group comprising D-E-F, in which

D is a valence bond or a group C_{1-6}alkylene, C_{1-6}alkenylene, C_{1-6}alkynylene, C_{5-7}arylene, or C_{5-7}heteroarylene, said group optionally being substituted with 1 to 4 group R^{11},

E is, when present, a valence bond, O, S, NR^4, or a group C_{1-6}alkylene, C_{1-6}alkenylene, C_{1-6}alkynylene, C_{5-7}arylene, or C_{5-7}heteroarylene, said group optionally being substituted with 1 to 4 group R^{11},

F is, when present, a valence bond, O, S, or NR^4,

A is a spacing group distancing the chemical structure from the complementing element, which may be a nucleic acid,

R^1, R^2, and R^3 are independent of each other selected among the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_4-C_8 alkadienyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R^4, 0-3 R^5 and 0-3 R^9 or C_1-C_3 alkylene-NR^4_2, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 alkylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9,

FEP is a group selected among the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_4-C_8 alkadienyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R^4, 0-3

R^5 and 0-3 R^9 or C_1-C_3 alkylene-NR^4_2, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 alkylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9,

where R^4 is H or selected independently among the group consisting of C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, heteroaryl, said group being substituted with 0-3 R^9 and

R^5 is selected independently from $-N_3$, -CNO, $-C(NOH)NH_2$, -NHOH, $-NHNHR^6$, $-C(O)R^6$, $-SnR^6_3$, $-B(OR^6)_2$, $-P(O)(OR^6)_2$ or the group consisting of C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_4-C_8 alkadienyl said group being substituted with 0-2 R^7,

where R^6 is selected independently from H, C_1-C_6 alkyl, C_3-C_7 cycloalkyl, aryl or C_1-C_6 alkylene-aryl substituted with 0-5 halogen atoms selected from -F, -Cl, -Br, and –I; and R^7 is independently selected from $-NO_2$, $-COOR^6$, $-COR^6$, -CN, $-OSiR^6_3$, $-OR^6$ and $-NR^6_2$.

R^8 is H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_7 cycloalkyl, aryl or C_1-C_6 alkylene-aryl substituted with 0-3 substituents independently selected from –F, -Cl, $-NO_2$, $-R^3$, $-OR^3$, $-SiR^3_3$

R^9 is =O, -F, -Cl, -Br, -I, -CN, $-NO_2$, $-OR^6$, $-NR^6_2$, $-NR^6-C(O)R^8$, $-NR^6-C(O)OR^8$, $-SR^6$, $-S(O)R^6$, $-S(O)_2R^6$, $-COOR^6$, $-C(O)NR^6_2$ and $-S(O)_2NR^6_2$.

In a preferred embodiment Z is O or S, P is a valence bond, Q is CH, B is CH_2, and R^1, R^2, and R^3 is H. The bond between the carbonyl group and Z is cleavable with aqueous I_2.

<u>Partitioning</u>

The partitioning step, by which the library of bifunctional molecules is subjected to a condition partitioning one or more complexes having a certain property from the remainder of the library, may be referred to as the enrichment step or the selection step, as appropriate, and includes the screening of the library for encoded molecules having predetermined desirable characteristics. Predetermined desirable characteristics can include binding to a target,

catalytically changing the target, chemically reacting with a target in a manner which alters/modifies the target or the functional activity of the target, and covalently attaching to the target as in a suicide inhibitor.

In theory, molecules of interest can be selected based on their properties using either physical or physiological procedures. The method preferred according to the present invention is to enrich molecules with respect to binding affinity towards a target of interest. In a certain embodiment, the basic steps involve mixing the library of complexes with the immobilized target of interest. The target can be attached to a column matrix or microtitre wells with direct immobilization or by means of antibody binding or other high-affinity interactions. In another embodiment, the target and displayed molecules interact without immobilisation of the target. Displayed molecules that bind to the target will be retained by a filter, size-exclusion chromatography *etc*, while non-binding displayed molecules will be removed during a single or a series of wash steps. The identifiers of complexes bound to the target can then be separated by cleaving a physical connection to the encoded molecule or the entire complex may be eluted. It may be considered advantageously to perform a chromatography step after or instead of the washing step. After the cleavage of the physical link between the synthetic molecule and the identifier, the identifier may be recovered from the media and optionally amplified before the decoding step.

A significant reduction in background binders may be obtained with increased washing volumes, repeating washing steps, higher detergent concentrations and prolonged incubation during washing. Thus, the more volume and number of steps used in the washing procedure together with more stringent conditions will more efficiently remove non-binders and background binders. The right stringency in the washing step can also be used to remove low-affinity specific binders. However, the washing step will also remove wanted binders if too harsh conditions are used.

A blocking step, such as incubation of solid phase with skimmed milk proteins or other inert proteins and/or mild detergent such as Tween-20 and Triton X-100, may also be used to reduce the background. The washing conditions should be as stringent as possible to remove background binding but to retain specific binders that interact with the immobilized target. Generally, washing conditions are adjusted to maintain the desired affinity binders, e.g. binders in the micromolar, nanomolar, or pocomolar range.

The target can be any compound of interest. E.g. the target can be a protein, peptide, carbohydrate, polysaccharide, glycoprotein, hormone, receptor, antigen, antibody, virus, substrate, metabolite, transition state analogue, cofactor, inhibitor, drug, dye, nutrient, growth factor, cell, tissue, etc. without limitation. Suitable targets include, but are not limited to, angiotensin converting enzyme, renin, cyclooxygenase, 5-lipoxygenase, IIL- 1 0 converting enzyme, cytokine receptors, PDGF receptor, type II inosine monophosphate dehydrogenase, β-lactamases, integrin, proteases like factor VIIa, kinases like Bcr-Abl/Her, phosphotases like PTP-1B, and fungal cytochrome P-450. Targets can include, but are not limited to, bradykinin, neutrophil elastase, the HIV proteins, including *tat, rev, gag, int*, RT, nucleocapsid etc., VEGF, bFGF, TGFβ, KGF, PDGF, GPCR, thrombin, substance P, IgE, sPLA2, red blood cells, glioblastomas, fibrin clots, PBMCs, hCG, lectins, selectins, cytokines, ICP4, complement proteins, etc.

A target can also be a surface of a non-biological origin, such as a polymer surface or a metal surface. The method of the invention may then be used to identify suitable coatings for such surfaces.

In a preferred embodiment, the desirable synthetic molecule acts on the target without any interaction between the nucleic acid attached to the desirable encoded molecule and the target. In one embodiment, the bound complex-target aggregate can be partitioned from unbound complexes by a number of methods. The methods include nitrocellulose filter binding, column chroma-

tography, filtration, affinity chromatography, centrifugation, and other well known methods. A preferred method is size-exclusion chromatography.

Briefly, the library of complexes is subjected to the target, which may include contact between the library and a column onto which the target is immobilised. Identifiers associated with undesirable encoded molecules, i.e. synthetic molecules not bound to the target under the stringency conditions used, will pass through the column. Additional undesirable encoded molecules (e.g. encoded molecules which cross-react with other targets) may be removed by counter-selection methods. Desirable complexes are bound to the column. The target may be immobilized in a number of ways. In one embodiment, the target is immobilized through a cleavable physical link, such as one more chemical bonds. The aggregate of the target and the complex may then be subjected to a size exclusion chromatography to separate the aggregate from the rest of the compounds in the media. The complex may then be eluted from the target by changing the conditions (e.g., salt, pH, surfactant, temperature etc.). Alternatively, the complex may be provided with a cleavable linker, preferable orthogonal to a cleavable linker that attaches the target to the solid support, at a position between the synthetic molecule and the identifier. Subsequent to the size exclusion chromatography this cleavable linker is cleaved to separate the identifiers of complexes having affinity towards the targets. Just to mention a single type of orthogonal cleavable linkages, one could attach the target to the solid support through a linkage that can be cleaved by a chemical agent, and the linker separating the synthetic molecule and the identifier may be selected as a photocleavable linkage. More specifically, the former linkage may be a disulphide bond that can be cleaved by a suitable reducing agent like DTT (dithiothreitol) and the latter linkage may be a *o*-nitrophenyl group.

There are other partitioning and screening processes which are compatible with this invention that are known to one of ordinary skill in the art. In one embodiment, the products can be fractionated by a number of common

methods and then each fraction is assayed for activity. The fractionization methods can include size, pH, hydrophobicity, etc.

Inherent in the present method is the selection of encoded molecules on the basis of a desired function; this can be extended to the selection of molecules with a desired function and specificity. Specificity can be required during the selection process by first extracting complexes which are capable of interacting with a non-desired "target" (negative selection, or counter-selection), followed by positive selection with the desired target. As an example, inhibitors of fungal cytochrome P-450 (fungicides) are known to cross-react to some extent with mammalian cytochrome P-450 (resulting in serious side effects). Highly specific inhibitors of the fungal cytochrome could be selected from a library by first removing those complexes capable of interacting with the mammalian cytochrome, followed by retention of the remaining products which are capable of interacting with the fungal cytochrome.

Brief Description of the Figures

Fig. 1 discloses two embodiments of using a Taqman probe (5' nuclease probe) in the measurement of the presence or absence of a certain codon.
Fig. 2 discloses a standard curve used in example 1.
Fig. 3 discloses the result of two experiments reported in example 1.

Example 1

A preferred embodiment of the invention utilizing a universal Taqman probe is shown in Fig.1. Four codon are shown (P1 through P4; bold pattern) along with flanking regions (light pattern). A universal Taqman probe anneals to a region adjacent to the codon region, but within the amplicon defined by the universal PCR primers Pr.1 and Pr. 2. These primers could be the same as used for amplification of the identifier oligonucleotides encoding binders after an enrichment process on a specific target. However, are minimal length templates preferred during the encoding process, the region involved in Taqman probe annealing could be appended to the library identifier oligonu-

cleotides by e.g. overlap PCR, ligation, or by employing a long downstream PCR primer containing the necessary sequences. The added length corresponding to the region necessary for annealing of the Taqman probe would be form 20 to 40 nts depending on the type of TaqMan probe and T_A of the PCR primers. The Q-PCR reactions are preferably performed in a 96- or 384-well format on a real-time PCR thermocycling machine.

Fig. 1A shows the detection of abundance of a specific codon sequence in position one. Similar primers are prepared for all codon sequences. For each codon sequence utilized to encode a specific BB in the library a Q-PCR reaction is performed with a primer oligonucleotide complementary to the codon sequence in question. A downstream universal reverse primer Pr. 2 is provided after the Taqman probe to provide for an exponential amplification of the PCR amplicon. The setup is most suited for cases where the codon constitutes a length corresponding to a length suitable for a PCR primer.

Fig. 1B shows the detection of abundance of a specific codon sequence in a specific codon position using a primer which is complementing a codon and a framing sequence. Similar primers are used for all the codons and framing sequences. For each codon sequence utilized to encode a specific BB at a specific codon position in the library a Q-PCR reaction is performed with an oligo complementary to the codon sequence in question as well as a short region up- or downstream of the codon region which ensures extension of the primer in a PCR reaction only when annealed to the codon sequence in that specific codon position. The number of specific primers and Q-PCR reactions needed to cover all codon sequences in all possible codon positions equals the number of codon sequences times the number of codon positions. Thus, monitoring the abundance of 96 different codon sequences in 4 different positions can be performed in a single run on four 96 wells micro titre plates (as shown in Fig. 1B) or a single 384 well plate on a suitable instrument. This architecture allows for the decoding of a 8,5 $^{*}10^{7}$ library of different encoded molecules.

Quantification is performed relative to the amount of full-length PCR product obtained in a parallel control reaction on the same input material performed with the two external PCR primers Pr.1 + Pr. 2. Theoretically, a similar rate of accumulation of this control amplicon compared to the accumulation of a product utilizing a single codon + sequence specific primer would indicate a 100% dominance of this particular sequence in the position in question.

Although the setups shown in Fig. 1A and 1B employ a Taqman probe strategy, other detection systems (SYBR green, Molecular Beacons etc.) could be utilized. In theory, multiplex reactions employing up to 4 different fluorofors in the same reaction could increase throughput correspondingly.

An example of how a deconvolution process of a library of encoded molecules occurs is described in the following. Imagine that at the end of a selection scheme a pool of 3 ligand families (and the corresponding coding templates) are dominating the population and present at approx. the same concentration. Three different chemical entities are present in the first position of the encoded compounds, and each of these chemical entities are present in combination with one unique chemical entity out of 3 different chemical entities in position P2. Only one chemical entity in position 3 gives rise to active binders, whereas any of a 20% subset of chemical entities (*e.g.* determined by charge, size or other characteristica) are present in position 4. The outcome of the initial codon profile analysis would be: 3 codon sequences are equally dominating in position P1, 3 other codon sequences in position P2, 1 unique codon sequence is dominant in P3 whereas somewhat similarly increased levels of 20% of the codon sequences (background levels of the remaining 80% sequences) are seen in P4. In such cases it could be relevant to use an iterative Q-PCR ("IQ-PCR") strategy to perform a further deconvolution of a library after selection. Again with reference to the example above, by taking the PCR products from the 3 individual wells that contained primers giving the high yields in position P1, diluting the product appropriately and

performing a second round of Q-PCR on each of these identifier oligonucleotides separately, it would be possible to deduce which codon sequence(s) is preferred in P2 when a given codon sequence is present in P1.

Experimental example: The 10 templates used for Q-PCR quantification

Taqman MGB probe binding region: *=AATTCCAGCTTCTAGGAAGAC

P1 P2 P3 P4

5'-CAGCTTGGACACCACGTCATACTAGCTGCTAGAGATGTGGTGATATTAGTGTGTGACGATGGTACGCACAAGTACGAACGTGCATCAGAGAGGACGAGCAGGACCTGGAACCTGGTGC*TTCCTCCACCACGTCTCTGAC-3'

GGAAGAAGACAGAAGACCTG	CTCGACCACTGCAGGTGGAGCTCC
TCAGGAGTCGAGAACTGAAG	CGTGCTTCCTCTGCTGCACCACCG
TGTGTACGTCAACACGTCAG	CCTGGTGTCGAGGTGAGCAGCAGC
TGTGGAACTACCATCCAAGG	CTCGACGAGGTCCATCCTGGTCGC
CCATCCAACATCGTTGGAAG	CGTGAGGAGCAGGTCCTCCTGTCG
AACCTGTCCTGTGAGATCTG	CCTGACACTGGTCGTGGTCGAGGC
TCACGAAGCTGGATGATGAG	CCATCTCGACGACCTGCTCCTGGG
TAGCATCGATCGAACGTAGG	CCACGAGGTCTCCACTGGTCCAGG
TCGAAGCTACTGTCGAGATG	CCACTGAGCTGCTCCTCCAGGTGG

Oligos for template synthesis:

FPv2: CAGCTTGGACACCACGTCATAC
RPv2: GTCAGAGACGTGGTGGAGGAA

Temp1-1: CAGCTTGGACACCACGTCATACTAGCTGCTAGAGATGTGGTGATATTAGTGTGTGACGAT
Temp1-2: CAGCTTGGACACCACGTCATACGGAAGAAGACAGAAGACCTGATATTAGTGTGTGACGAT
Temp1-3: CAGCTTGGACACCACGTCATACTCAGGAGTCGAGAACTGAAGATATTAGTGTGTGACGAT
Temp1-4: CAGCTTGGACACCACGTCATACTGTGTACGTCAACACGTCAGATATTAGTGTGTGACGAT
Temp1-5: CAGCTTGGACACCACGTCATACTGTGGAACTACCATCCAAGGATATTAGTGTGTGACGAT
Temp1-6: CAGCTTGGACACCACGTCATACCCATCCAACATCGTTGGAAGATATTAGTGTGTGACGAT
Temp1-7: CAGCTTGGACACCACGTCATACAACCTGTCCTGTGAGATCTGATATTAGTGTGTGACGAT
Temp1-8: CAGCTTGGACACCACGTCATACTCACGAAGCTGGATGATGAGATATTAGTGTGTGACGAT
Temp1-9: CAGCTTGGACACCACGTCATACTAGCATCGATCGAACGTAGGATATTAGTGTGTGACGAT
Temp1-10: CAGCTTGGACACCACGTCATACTCGAAGCTACTGTCGAGATGATATTAGTGTGTGACGAT

Temp2: GTCCTCTCTGATGCACGTTCGTACTTGTGCGTACCATCGTCACACACTAATATC

Temp3-1: GAACGTGCATCAGAGAGGACGAGCAGGACCTGGAACCTGGTGCAATTCCAGCTTCTAGGAAGACT
Temp3-2: GAACGTGCATCAGAGAGGACTCGACCACTGCAGGTGGAGCTCCAATTCCAGCTTCTAGGAAGACT
Temp3-3: GAACGTGCATCAGAGAGGACGTGCTTCCTCTGCTGCACCACCGAATTCCAGCTTCTAGGAAGACT
Temp3-4: GAACGTGCATCAGAGAGGACCTGGTGTCGAGGTGAGCAGCAGCAATTCCAGCTTCTAGGAAGACT
Temp3-5: GAACGTGCATCAGAGAGGACTCGACGAGGTCCATCCTGGTCGCAATTCCAGCTTCTAGGAAGACT
Temp3-6: GAACGTGCATCAGAGAGGACGTGAGGAGCAGGTCCTCCTGTCGAATTCCAGCTTCTAGGAAGACT
Temp3-7: GAACGTGCATCAGAGAGGACCTGACACTGGTCGTGGTCGAGGCAATTCCAGCTTCTAGGAAGACT
Temp3-8: GAACGTGCATCAGAGAGGACCATCTCGACGACCTGCTCCTGGGAATTCCAGCTTCTAGGAAGACT
Temp3-9: GAACGTGCATCAGAGAGGACCACGAGGTCTCCACTGGTCCAGGAATTCCAGCTTCTAGGAAGACT
Temp3-10: GAACGTGCATCAGAGAGGACCACTGAGCTGCTCCTCCAGGTGGAATTCCAGCTTCTAGGAAGACT

Temp4: GTCAGAGACGTGGTGGAGGAAGTCTTCCTAGAAGCTGGAATT

Synthesis of identifier oligonucleotides:

The 10 identifier oligonucleotides were assembled in 10 seperate 50 µl PCR reactions each containing 0.05 pmol of the oligos Q-Temp1-X, Q-Temp2, Q-Temp3-X and Q-Temp4 (x=1 through 10) and 25 pmol of the external primers FPv2 and RPv2 with TA= 53°C. The 160 bp products were gel-purified using QIAquick Gel Extraction Kit from QIAGEN (Cat. No. 28706) and quantified on spectrophotometer. As a control, 20 ng of each of the templates (as estimated from these measurements) were loaded on an agarose gel.

Preparation of samples for Q-PCR:

Sample A: Generated by mixing 20 ng from each identifier oligonucleotide prep. Volume was adjusted to 50 µl. Concentration: 4 ng/µl = 38.46 fmol/µl (160bp x 650 Da/bp =1.04x105 g/mol. 1 ng= 9.615 fmol). Diluted to 10^7 copies/5µl (0.00332 fmol/µl).

Sample B: 20 ng/20µl stocks of each template were prepared. The sample was mixed as follows:

5µl undil. Template #10
5µl 2x dil. Template #9
5µl 4x dil. Template #8
5µl 8x dil. Template #7
5µl 16x dil. Template #6
5µl 32x dil. Template #5
5µl 64x dil. Template #4
5µl 128x dil. Template #3
5µl 256x dil. Template #2
5µl 512x dil. Template #1

Concentration: 10ng/50µl= 0.20 ng/µl = 1.923 fmol/µl. Diluted 579.2-fold to 10^7 copies/5µl (0.00332 fmol/µl).

Standard curve: The samples for the standard curve was prepared by diluting Sample A 116.55-fold to 10^9 copies/5 µl (0.33 fmol//µl) and subsequently per-

forming a 10-fold serial dilution of this sample. 5 µl was used for each PCR reaction. The standard curve is shown in Fig. 2.

Q-PCR reactions

For 5 ml premix (for one 96-well plate):
2.5 ml Taqman Universal PCR Master Mix (Applied Biosystems; includes Taq polymerase, dNTPs and optimized Taq pol. buffer)
450 µl RPv2 (10 pmol/ul)
25 µl Taqman probe (6-FAM-TCCAGCTTCTAGGAAGAC-MGBNFQ; 50 µM; Applied Biosystems)
1075 µl H2O

40.5 µl premix was aliquoted into each well and 4.5 µl of relevant upstream PCR primer (FPv2 (for standard curve) or one of the codon specific primers listed below; 10 pmol/µl) and 5 µl sample (H2O in wells for negative controls) was added. The codon-specific PCR primers were: (Tm calculations shown are from Vector NTI; matched to Tm for RPv2 (67.7°C))

P1-1:	GTCATACTAGCTGCTAGAGATGTGGTGATA	66.8°C
P1-2:	CATACGGAAGAAGACAGAAGACCTGATA	67.8°C
P1-3:	TCATACTCAGGAGTCGAGAACTGAAGATA	67.6°C
P1-4:	CATACTGTGTACGTCAACACGTCAGATA	67.4°C
P1-5:	CATACTGTGGAACTACCATCCAAGGATA	68.0°C
P1-6:	CCATCCAACATCGTTGGAAGAT	67.8°C
P1-7:	CATACAACCTGTCCTGTGAGATCTGATA	67.7°C
P1-8:	ATACTCACGAAGCTGGATGATGAGATA	67.3°C
P1-9:	CATACTAGCATCGATCGAACGTAGGATA	68.1°C
P1-10:	TCATACTCGAAGCTACTGTCGAGATGATA	68.2°C
P2-1:	ATATTAGTGTGTGACGATGGTACGCA	67.8°C
P3-1:	ACAAGTACGAACGTGCATCAGAGA	67.7°C
P4-1:	CGAGCAGGACCTGGAACCT	67.7°C
P4-2:	TCGACCACTGCAGGTGGA	68.3°C

P4-3:	GCTTCCTCTGCTGCACCA	66.7°C
P4-4:	GGTGTCGAGGTGAGCAGCA	69.1°C
P4-5:	CGACGAGGTCCATCCTGGT	68.6°C
P4-6:	GTGAGGAGCAGGTCCTCCTGT	68.0°C
P4-7:	CTGACACTGGTCGTGGTCGA	68.8°C
P4-8:	CATCTCGACGACCTGCTCCT	67.9°C
P4-9:	ACGAGGTCTCCACTGGTCCA	68.3°C
P4-10:	ACTGAGCTGCTCCTCCAGGT	66.5°C

Thermocycling/measurement of fluoresence was performed on an Applied Biosystems ABI Prism 7900HT real-time instrument utilizing the standard cycling parameters:

95°C 10 min;

40 cycles of

95°C 15 sec;

60°C 1 min

All samples were run in duplicate.

Results

Fig. 2 shows the standard curve calculated by the 7900HT system software. The log of the starting copy number was plotted against the measured C_T value. The relationship between C_T and starting copy number was linear in the range from 10 to 10^9 template copies.

This standard curve was utilized by the system software to calculate the quantity in the "unknown" samples as shown below.

Table I: Sample A (Shown graphically in Fig. 3A

Sample A: Equi-molar ratios	Observed A	Observed B	Expected
FPv2	12539947,00	11977503,00	10000000
P1-1	445841,90	480382,03	1000000
P1-2	884840,70	847478,56	1000000
P1-3	1013073,56	948770,00	1000000
P1-4	764187,94	741304,40	1000000
P1-5	1352874,60	1275155,50	1000000
P1-6	1284075,60	1337928,50	1000000
P1-7	658161,80	747371,56	1000000
P1-8	742187,20	653874,00	1000000
P1-9	824587,75	705785,75	1000000
P1-10	813550,75	836037,90	1000000
P2-1	13145159,00	14482606,00	10000000
P3-1	13263911,00	12773780,00	10000000
P4-1	1430704,80	1472576,80	1000000
P4-2	2681652,00	2481824,80	1000000
P4-3	1933106,80	2085476,40	1000000
P4-4	1359684,40	1364621,40	1000000
P4-5	2206709,80	2065813,60	1000000
P4-6	1652718,10	1873777,20	1000000
P4-7	1468208,10	1416153,00	1000000
P4-8	1664467,50	1581067,00	1000000
P4-9	1462520,60	1594593,80	1000000
P4-10	2020088,20	1912277,40	1000000

Table II: Sample B (Shown graphically in Fig. 3B)

Sample B: 2-fold	Observed A	Observed B	Expected

dil.

FPv2	4,97E+06	5,05E+06	10000000
P1-1	9955,07	10899,97	9765,625
P1-2	12732,32	13469,12	19531,25
P1-3	25542,8	25419,85	39062,5
P1-4	34748,89	44070,81	78125
P1-5	110881,41	123734,13	156250
P1-6	163687,44	166220,5	312500
P1-7	156993,81	172005,64	625000
P1-8	343176,78	374809,13	1250000
P1-9	646619,44	576151	2500000
P1-10	1,49E+06	1,72E+06	5000000
P2-1	5,19E+06	5,37E+06	10000000
P3-1	5,29E+06	5,09E+06	10000000
P4-1	(no signal)	70223,8	9765,625
P4-2	42103,32	22733,17	19531,25
P4-3	54480,62	39663,62	39062,5
P4-4	51293,07	43950,9	78125
P4-5	137946,95	115027,34	156250
P4-6	174134,64	156442,55	312500
P4-7	316505,78	283856,84	625000
P4-8	737661,44	691296,75	1250000
P4-9	1,42E+06	1,45E+06	2500000
P4-10	3,72E+06	3,52E+06	5000000

The results of the experiments show the possibility of accurately quantification of identifier oligonucleotides down to or even below 10 copies with a 9 fold dynamic range, and reliable relative quantification of the tested codons in various positions in the identifier oligonucleotide.

While the invention has been described with references to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention. All patent and literature references cited herein are hereby incorporated by reference in their entirety.

Claims

1. A method for obtaining structural information about an encoded molecule produced by a process comprising reaction of a plurality of chemical entities, said encoded molecule being capable of forming part of a complex also comprising an identifier oligonucleotide containing codons informative of the identity of chemical entities which have participated in the formation of the encoded molecule, the method comprises

   a) mixing a primer oligonucleotide with the identifier oligonucleotide,

   b) subjecting the mixture to a condition allowing for an extension reaction to occur when the primer is sufficient complementary to a part of the identifier oligonucleotide, and

   c) evaluating, based on measurement of the extension reaction, the presence, absence, or relative abundance of one or more codons.

2. The method according to claim 1, wherein a composition of one, two, or more identifier oligonucleotides are processed simultaneously.

3. The method according to claim 2, the composition is a result of subjecting a library of different complexes to a condition partitioning one or more complexes having a certain property from the remainder of the library.

4. The method according to claim 1, wherein the condition which allows for an extension reaction to occur includes a polymerase or a ligase as well as suitable substrates.

5. The method according to claim 4, wherein the condition includes a polymerase and a substrate comprising a blend of (deoxy)ribonucleotide triphosphates.

6. The method according to any of the preceding claims, wherein the chemical entities are precursors for structural units appearing in the encoded molecule.

7. The method according to any of the claims 1 to 6, wherein the process of producing the one or more encoded molecules comprises transferring one

or more chemical entities to a nascent encoded molecule by a building block which further comprises an anti-codon.

8. The method of claim 7, wherein the information of the anti-codon is transferred in conjunction with the chemical entity to the nascent encoded molecule.

9. The method according to any of the preceding claims 1 to 8, wherein the identifier comprises two or more codons.

10. The method according to any of the preceding claims 1 to 8, wherein the identifier comprises three or more codons.

11. The method according to any of the preceding claims, wherein neighbouring codons of the identifier are spaced by a framing sequence.

12. The method according to claim 11, wherein the framing sequence positions the reaction of a chemical entity in the synthesis history of the encoded molecule.

13. The method according to any of the claims 1 to 12, wherein at least a part of the primer is complementary to a codon.

14. The method according to claims 1 to 13, wherein at least a part of the primer is complementary to a codon and an adjacent framing sequnece.

15. The method according to any of the claims 1 to 13, wherein the codons have a length of four or more nucleotides.

16. The method according to any of the claims 1 to 15, wherein the sequence comprising the codon and an adjacent framing sequence has a total length of 11 nucleotides or more.

17. The method according to any of the claims 1 to 16, wherein the extension reaction is measured using the polymerase chain reaction (PCR), wherein the primer of claim 1 is involved in said PCR.

18. The method according to any of the claims 1 to 17, wherein a primer is labelled.

19. The method according to claim 18, wherein the primer is labelled with a small molecule, a radio active component, or a fluorogenic molecule.

20. The method according to claim 19, wherein the small molecule label is selected from biotin, dinitrophenol, and digoxigenin, and the PCR amplicons

are detected using an enzyme labelled streptavidin, anti-dinitrophenol. or anti-digoxigenin, respectively, reporter molecule.

21. The method according to any of the claims 1 to 19, wherein extension reaction is measured by real-time PCR.

22. The method according to claim 21, wherein the real-time PCR involves the use of an oligonucleotide probe responsible for the generation of a detectable signal during the propagation of the PCR reaction.

23. The method according to any of the claims 1 to 21, wherein the probe is designed to hybridise at a position downstream of a primer binding site.

24. The method according to claim 22 or 23, wherein the probe is a 5' nuclease oligoprobe or a hairpin oligoprobe.

25. The method according to claims 2 or 3, wherein the library comprises complexes with identifier oligonucleotides having n codon positions and the codons in said codon positions being selected from a set of m different codons.

26. The method according to claim 25, wherein a framing sequence is related to each of the n codon positions in a particular complex, said framing sequence positions the reaction of a chemical entity in the synthesis history of the encoded molecule.

27. The method according to claim 25, wherein each codon in the set of m different codons differs from any other codons in the set in at least two nucleotide positions.

28. The method according to claim 26, wherein each framing sequence in a set of n different framing sequences differs from any other framing sequences in the set in at least two nucleotide positions.

29. A method for identifying the chemical entities utilized in the formation of an encoded molecule or a composition of encoded molecules, wherein in separate compartments, n x m primers individually are mixed with an aliquot of a composition obtained by subjecting a library of different complexes to a condition partitioning said composition from the remainder of the library, subjected to a mixture of polymerase and substrate (deoxy)ribonucleotide triphosphates under conditions allowing for an extension reaction to occur

when a primer is sufficient complementary to a part of one or more identifier oligonucleotides present in the aliquot, and evaluation, based on measurement of the extension reaction, the presence, absence, or relative abundance of one or more codons in each compartment.

30. A set comprising a collection of oligonucleotide primers, a polymerase, a composition of (deoxy)ribonucleotide triphosphates (dNTPs), and a library of complexes composed of a display molecule part and an identifier oligonucleotide, said oligonucleotide comprising codons informative of the identity of the chemical entities which has participated in the formation of the display molecule, wherein the oligonucleotide primers are sufficient complementary to codons appearing on the identifier oligo nucleotides in the library to allow for an extension reaction to occur.

31. An encode molecule identified by a method according to claims 1 to 28.

Abstract

The present invention relates to a method for obtaining structural information about an encoded molecule. The encoded molecule may be produced by a reaction of a plurality of chemical entities and may be capable of being connected to an identifier oligonucleotide containing codons informative of the identity of the chemical entities which have participated in the formation of the encoded molecule. In a certain embodiment, primers are designed complementary to the codons appearing on the identifier oligonucleotide, and the presence, absence or relative abundance of a codon is evaluated by mixing a primer with the identifier oligonucleotide in the presence of a polymerase and substrate (deoxy)ribonucleotide triphosphates and measuring the extension reaction.

Fig. 1A

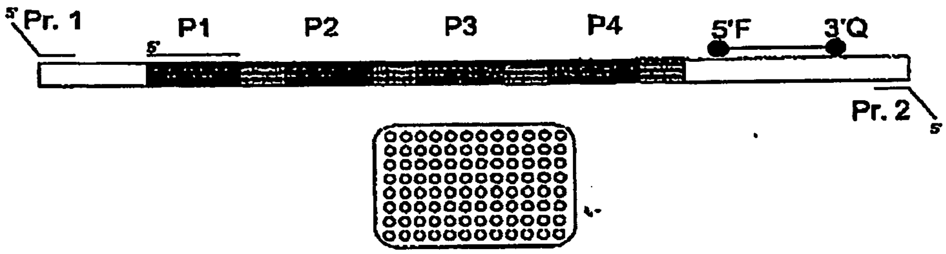
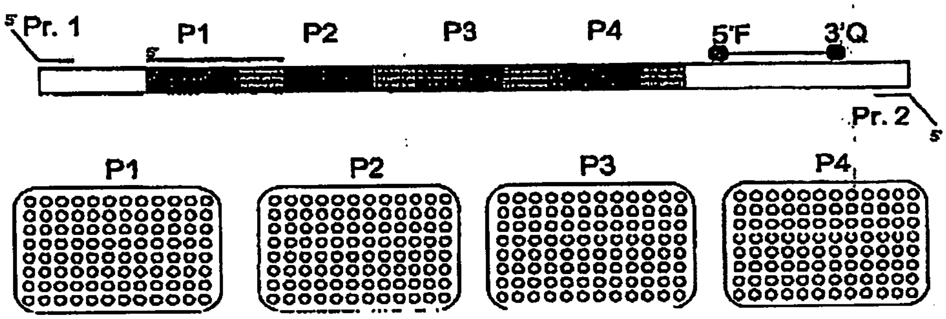

5' Pr. 1
P1
5'
P2
P3
P4
5'F
3'Q
Pr. 2
5'

Fig. 1B


5' Pr. 1
P1
5'
P2
P3
P4
5'F
3'Q
Pr. 2
5'
P1
P2
P3
P4

Modtaget

18 SEP. 2003

PVS

2

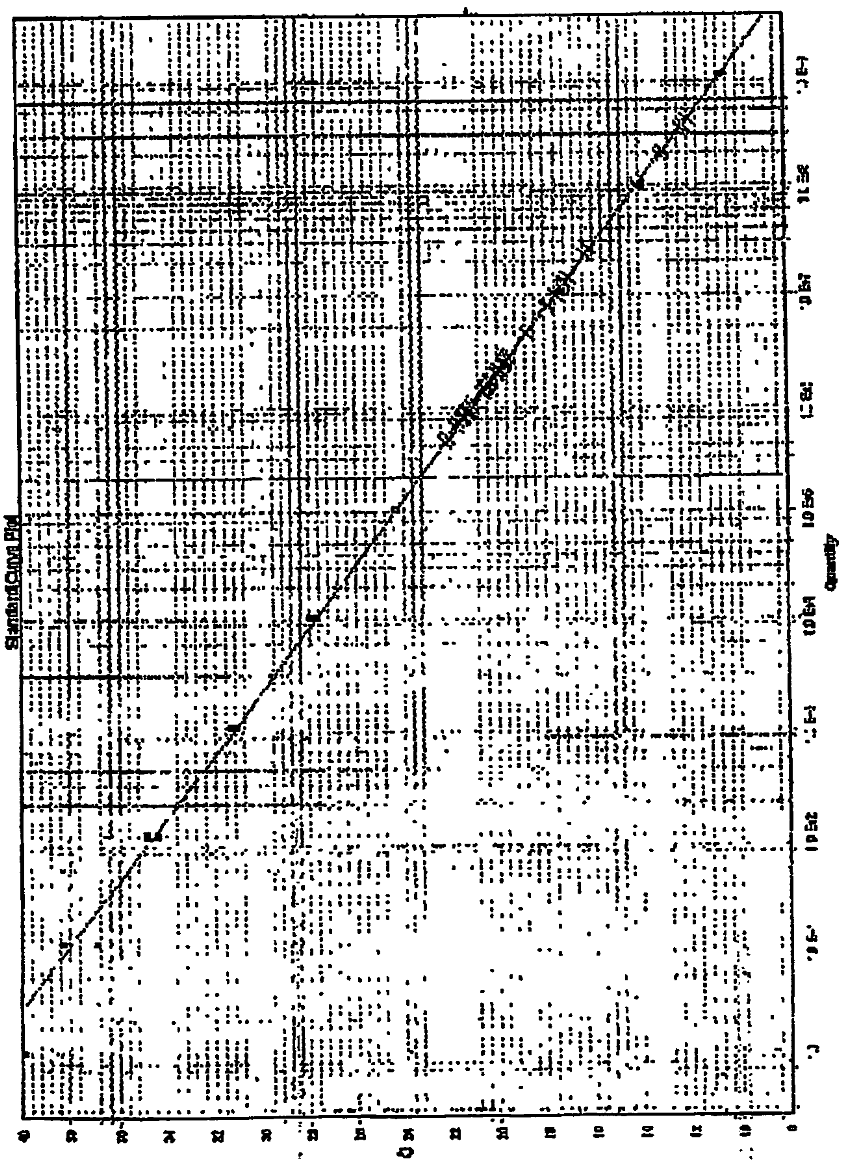


Fig 2

Modtaget

18 SEP. 2003

PVS

Fig 3A

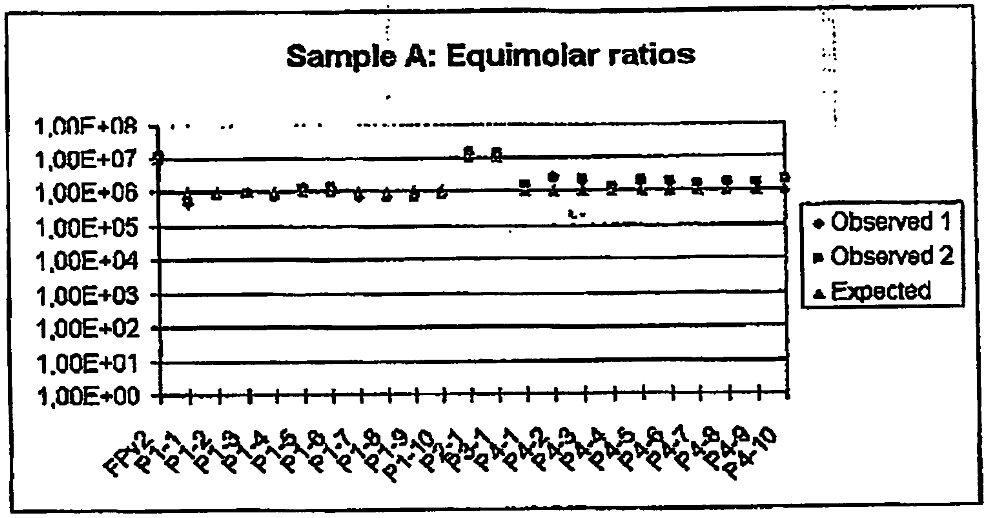


Fig 3B **Sample B**